Math 340 **Matt Advised Problems: MAPs™** Midterm 2

This exam covers the material discussed in lecture from Chapters 4, 6 of our book and Chapter 6 of the other book we used for CLT stuff. The best way to study is to (listed in order of importance) understand the 4 problems here, the homework problems, the examples presented in lecture, and the extra problems provided from the book.

Topics NOT covered:

- Proofs
- Stuff from Midterm 1
- Chebyshev's inequality
- Probability generating functions: $f(t) = Et^X$.
- Conditional distributions

Topics covered:

1. Mean and variance
   - EX and $Er(X) = \sum r(x)P(X = x)$
   - Calculating and interpreting var X and $\sigma(X) = \sqrt{\text{var}(X)}$.
   - $T = Y + (1 - Y)(T' + T'' + 1)$ equation for time it takes to reach 1 from HW4
   - Sums of random variables
     - sum of binomials, poissons, geometrics
     - expectation and variance of sums
     - coupon collector
   - Law of large numbers
2. Central Limit Theorem
   - Using the table
   - Using CLT
   - Histogram correction
   - confidence intervals
   - sample size
3. Conditional Probability
   - definition
   - two stage experiments
   - baye's formula
   - conditional expectation

Additional book questions:

Section 4.6 9, 25, 29, 32, 40, 48

Section 6.7 (other book) 4, 38, 41, 50

Section 6.5 10, 26, 31, 39, 58

Additional questions:

1. Alice and Bob are sharing a huge pile of french fries and deciding how many to take by flipping coins. Alice's coin has probability 1/2 of heads. She eats two fries each time she flips a head. Bob's coin has probability 1/10 of landing heads. He eats five fries each time his coin lands heads. Let X be the number of flips until Alice eats 100 fries, and let Y be the number of flips until Bob does.

   (a) Find the mean and variance of X and Y.

   Solution: Notice we can write $X = \sum_{i=1}^{50} X_i$ with $X_i = \text{Geo}(1/2)$ and $Y = \sum_{i=1}^{20} Y_i$ with $Y_i = \text{Geo}(1/5)$. It follows that
   $$EX = 100, \quad EY = 200.$$
   $$\text{var}(X) = 50\,\text{var}(X_1) = \frac{1-\frac{1}{2}}{(1/2)^2} = 100.$$
   $$\text{var}(Y) = 20\frac{1-\frac{1}{10}}{(1/10)^2} = 1800.$$

   (b) Use normal approximation to estimate $P(\{X < 120\} \cap \{Y > 120\})$.

   Solution: Let's assume that $X = N(100, 10)$ so $P(X < 120) = .975$. We can set $Y = N(200, \sqrt{2250}) \approx N(200, 42)$. So $P(Y > 120) = .97$. The intersection has probability
   $$.975 * .97 = .945$$

   (c) Use the previous part to estimate $P(X < Y)$. Is your answer an over or underestimate? Explain how you could improve the estimate.

   Solution: The answer from the previous part is an estimate on this event. It is an underestimate though, because $\{X < 120\} \cap \{Y > 120\} \subseteq \{X < Y\}$. This could be improved by considering the event $E_k = \{X < k\} \cap \{Y > k\}$ and finding k that maximizes $P(E_k)$. We just figured out the case $k = 120$.

2. You are walking towards your favorite tree and start out one mile away. You move closer by first rolling a six-sided die to obtain the number D_i. You then cover a $1 - \frac{D_i}{10}$ fraction of the distance remaining between you and the tree. So if you roll a 3 on your first roll you would move to be .3 miles from the tree. If you then roll a 6 you would move to $.3 * .6$ miles away from the tree, and so on.

   (a) Write an expression for your distance, S_n, to the tree after n steps.

   Solution: $S_n = (D_1/10)\cdots(D_n/10)$.

   (b) What is ES_n?

   Solution: $ES_n = (3.5/10)^n = .35^n$

   (c) What is $E\ln(D_1/10)$?

   Solution: $\frac{1}{6}\sum_{k=1}^{6} \ln(k/10)$

(d) Explain why the law of large numbers guarantees $\ln S_n \approx nE\ln(D_1/10)$.

Solution: $\ln S_n = \sum_{i=1}^n \ln(D_i/10)$. This is a sum of iid random variables, so if we divide by n it ought to converge to $E \ln D_1$.

(e) What does $e^{\ln S_n}$ converge to?

Solution: $e^{(n/6)\sum \ln(k/10)} = ((.1 * 2. * 3 * .4. * 5 * .6)^{1/6})^n = .29^n$.

(f) There is a conflict between (b) and (e). Why are they different, and which answer actually predicts how close you will be to the tree after n steps?

Solution: Unlikely events make the expected value larger than what the most likely behavior will be. The answer to (e) is more accurate.

3. Let $X = \text{Geo}(p)$.

(a) What is $P(X = k)$?

Solution: $(1-p)^{k-1}p$

(b) What is $P(X = k + j \mid X > j)$?

Solution: We use the definition of $P(X = k + j \mid X \geq j)$:

$$\frac{P(\{X = k+j\} \cap \{X > j\})}{P(X > j)} = \frac{P(X = k+j)}{(1-p)^j} = \frac{(1-p)^{k+j-1}p}{(1-p)^j} = (1-p)^{k-1}p.$$

(c) The similarity between the previous two answers is referred to as the "memoryless property." Write in words why this is a fitting description.

Solution: This says that even if we know the wait time is at least j, the remaining time still has the same distribution. So there is no memory of the past time spent waiting.

(d) Let $X = \text{Geo}(1/3)$ and $Y = \text{Geo}(p)$. Find p so that $P(X \leq Y) = \frac{2}{3}$. Explain where you use the memoryless property.

Solution: Set $q = P(X \leq Y)$. Let $A_1 = \{X = 1\}$, $A_2 = \{X > 1\}$. We can use the recursion from Example 3.17 to write

$$q = 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3}q + q\frac{2}{3}(1-p).$$

The memoryless property is in the last term. We are assuming both X and Y are greater than 1. Set $q = 2/3$ and we have

$$\frac{2}{3} = \frac{1}{3} + \frac{2}{3}\frac{2}{3}(1-p)$$

This gives $p = 1/4$.

4. Ash (short for Ashley) is playing Pokémon Go. There are 151 different pokémon. In a special area all pokémon are equally likely to be encountered with the rule that: **the same pokémon is never**

encountered twice in a row. For example, if Pikachu was just encountered, then Pikachu will not appear in your next encounter. Instead, you are equally likely to meet any of the other **150 pokémon**.

(a) Suppose Ash has encountered i distinct pokémon, and Y_i is the number of pokémon she encounters until she has seen $i+1$ distinct pokémon. Explain why Y_i is a geometric random variable.

> **Solution:** Each box opening is an independent trial with the same probability of successfully finding a new coupon.

(b) What is the parameter for Y_0 and Y_i for $1 \leq i \leq 150$?

> **Solution:** 1, and $\frac{151-i}{150}$

(c) Let N be the (random) number of encounters needed for Ash to go from seeing 0 to all 151 pokémon. Write a formula for N in terms of the Y_i.

> **Solution:** $N = X_0 + X_1 + \cdots + X_{151}$.

(d) What is the expected value of N? It is okay to leave your answer as a sum.

> **Solution:** $1 + \sum_{i=1}^{n} \frac{150}{151-i}$.

(e) Suppose that the "never twice in a row" rule isn't in place. So, each encounter is equally likely to be any of the 151 pokémon. Let M be the number of encounters needed to see all 151 pokémon.

Without doing any calculations, explain which ought to be larger between EM and EN. What about var(M) and var(N)?

> **Solution:** $EM > EN$. This is because we are more likely to have repeats when we have a chance of encountering the pokémon last seen. var(M) > var(N) because the variance of a geometric increases as we decrease the success probability.